501.43507X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): KAIYA, et al

Serial No.: 10/774,470

Filed: February 10, 2004

Title: EXTERNAL STORAGE AND RECOVERY METHOD FOR EXTERNAL STORAGE AS WELL AS PROGRAM

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 9, 2004

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on:

Japanese Patent Application No. 2003-076865
Filed: March 20, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No.: 29,621

CIB/rr
Attachment

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: ２００３年　３月２０日

出 願 番 号
Application Number: 特願２００３－０７６８６５

[ST. 10/C]: ［ＪＰ２００３－０７６８６５］

出 願 人
Applicant(s): 株式会社日立製作所

２００４年　２月１９日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００４－３０１１０５３

【書類名】　特許願
【整理番号】　340201769
【あて先】　特許庁長官殿
【国際特許分類】　G06F 03/06
【発明者】
　【住所又は居所】　神奈川県小田原市中里３３２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
　【氏名】　海谷　佳一
【発明者】
　【住所又は居所】　神奈川県小田原市中里３３２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
　【氏名】　坪木　雅直
【発明者】
　【住所又は居所】　神奈川県小田原市中里３３２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
　【氏名】　水主　和人
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社日立製作所
【代理人】
　【識別番号】　100095371
　【弁理士】
　【氏名又は名称】　上村　輝之
【選任した代理人】
　【識別番号】　100089277
　【弁理士】
　【氏名又は名称】　宮川　長夫

【選任した代理人】

【識別番号】　100104891

【弁理士】

【氏名又は名称】　中村　猛

【手数料の表示】

【予納台帳番号】　043557

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　0110323

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラム

【特許請求の範囲】

【請求項１】　ホストコンピュータに接続される外部記憶装置であって、

前記ホストコンピュータにより利用されるデータを記憶する記憶手段と、

前記記憶手段を制御する制御手段とを有し、

前記制御手段は、

前記記憶手段に記憶されたデータに関して前記ホストコンピュータにより設定される回復可能時点を登録する登録手段と、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、

を備えたことを特徴とする外部記憶装置。

【請求項２】

前記登録手段は、前記ホストコンピュータにより設定される任意の複数時点を、前記回復可能時点として登録可能である請求項１に記載の外部記憶装置。

【請求項３】

前記記憶手段は、前記ホストコンピュータからの書込みデータをジャーナルデータとして記憶するジャーナルデータ記憶手段を有し、

前記登録手段は、前記ホストコンピュータからの指示に基づいて、前記ジャーナルデータの所定位置に標識情報を対応付けることにより、前記回復可能時点を登録するものである請求項１に記載の外部記憶装置。

【請求項４】

前記ジャーナルデータは、少なくとも、書込みデータと、書込み位置と、前記標識情報としての回復フラグ情報とを含んで構成され、

前記登録手段は、前記ジャーナルデータ中の所定の回復フラグ情報をセットすることにより、前記回復可能時点を登録するものである請求項３に記載の外部記

憶装置。

【請求項５】

前記記憶手段は、バックアップデータを記憶するバックアップデータ記憶手段を有し、

前記制御手段は、ジャーナルデータ管理手段を有し、

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶手段の空き容量が不足した場合には、前記ジャーナルデータ記憶手段に記憶されている最古のジャーナルデータを前記バックアップデータ記憶手段に移し替えて、前記ジャーナルデータ記憶手段の空き容量を増加させ、かつ、前記登録された回復可能時点のうち最古の回復可能時点が変更された旨を前記ホストコンピュータに通知するものである請求項３に記載の外部記憶装置。

【請求項６】

前記制御手段は、ジャーナルデータ管理手段を有し、

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶手段の空き容量が不足した場合には、前記記憶手段内の未使用の記憶領域を利用してジャーナルデータ記憶手段の論理サイズを自動的に拡張させるものである請求項３に記載の外部記憶装置。

【請求項７】　ホストコンピュータに接続された外部記憶装置のデータを、該外部記憶装置内で回復させるデータ回復方法であって、

記憶されたデータに関して前記ホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録ステップと、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する一覧送信ステップと、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復ステップと、

を含んだことを特徴とする外部記憶装置のデータ回復方法。

【請求項８】　ホストコンピュータに接続された外部記憶装置を制御するためのプログラムであって、

前記外部記憶装置は、前記ホストコンピュータにより利用されるデータを記憶

する記憶手段を有し、

前記記憶手段に記憶されたデータに関して前記ホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録手段と、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、

を、外部記憶装置のコンピュータ上に実現させるプログラム。

【請求項９】

ジャーナルデータを取得して前記記憶手段のジャーナルデータ記憶領域に記憶させるジャーナルデータ管理手段を、前記外部記憶装置のコンピュータ上に実現させると共に、

前記登録手段は、前記ホストコンピュータからの指示に基づいて、前記ジャーナルデータの所定位置に標識情報を対応付けることにより、前記回復可能時点を登録するものである請求項８に記載のプログラム。

【請求項１０】

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶領域の空き容量が不足した場合には、前記ジャーナルデータ記憶領域に記憶されている最古のジャーナルデータを前記記憶手段のバックアップデータ記憶領域に移し替えて、前記ジャーナルデータ記憶領域の空き容量を増加させ、前記登録された回復可能時点のうち最古の回復可能時点が変更された旨を前記ホストコンピュータに通知するものである請求項９に記載のプログラム。

【請求項１１】　外部記憶装置を利用するホストコンピュータを制御するプログラムであって、

前記外部記憶装置に記憶されたデータに関して、任意の複数時点で設定可能な回復可能時点を前記外部記憶装置に指示し登録させる登録指示手段と、

前記外部記憶装置に登録された前記回復可能時点の選択用情報を要求する選択用情報要求手段と、

前記外部記憶装置から受信した前記選択用情報に基づいて、所望のデータを所

望の回復可能時点まで回復させるように前記外部記憶装置に指示する回復指示手段と、

をホストコンピュータ上で実現させるためのプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、例えば、ディスク装置等の外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラムに関する。

【０００２】

【従来の技術】

比較的多量のデータを取り扱う業務用アプリケーションプログラム（データベースシステム）では、ホストコンピュータとは別体に形成されたディスクアレイ装置にデータを保存する。そして、ホストコンピュータのデータベースシステムからディスクアレイ装置上のデータにアクセスして、種々のデータ操作を行うようになっている。ディスクアレイ装置とは、複数のディスク装置をアレイ状に配設してなるもので、ホストコンピュータからの書込み命令や読出し命令等に応じて作動するようになっている。

【０００３】

ここで、データベースシステムが稼働中に、例えば、予期せぬ電源切断、オペレータの操作ミス、ハードウェア回路や他のプログラムの不調等によって、障害が発生した場合、データベースの内容を障害発生前の状態に復旧させる必要を生じる。また、障害以外に所望の時点までデータ操作を戻したい場合もある。

【０００４】

第１の従来技術として、通常のデータベースシステムでは、ホストコンピュータ上のデータベースシステム自身が、実データとは別にジャーナルデータ（ログデータ）をディスクアレイ装置の所定のディスク装置に書き出すようになっている。従って、通常のデータベースシステムでは、事前に取得してあるバックアップデータを基に、データベースシステム自身がディスク装置からジャーナルデータを読み出して、バックアップデータに順次反映させていく。これにより、ホス

トコンピュータ上のデータベースシステムは、ジャーナルデータが残存している範囲内で、所望の時点にデータベースを復旧させることができる。

【０００５】

第２の従来技術では、第１のディスク装置の内容を所定の周期でバックアップ用ディスク装置に保存すると共に、ジャーナルデータをジャーナル用ディスク装置に保存する。第１のディスク装置に障害が発生した場合は、バックアップデータ及びジャーナルデータに基づいて、第２のディスク装置内に仮想的な第１のディスク装置を生成し、第１のディスク装置へのデータアクセスを、仮想的な第１のディスク装置に内部的に切り替える。そして、第１のディスク装置の修復が完了すると、仮想的な第１のディスク装置の内容を第１のディスク装置に移し替えるようになっている（例えば、特許文献１参照）。

【０００６】

【特許文献１】

特開平６－１１０６１８号公報

【０００７】

【発明が解決しようとする課題】

上記第１の従来技術では、ホストコンピュータ上のデータベースシステム自身がジャーナルデータを管理しており、任意の時点にデータを復旧可能である。しかし、データベースシステム自身がデータ復旧作業を行うため、ホストコンピュータのコンピュータ資源（演算ユニットやメモリ等）が、データ復旧処理に使用されてしまい、復旧作業中に、本来の業務処理や他の業務処理の処理効率を低下させることになる。また、データベースシステムがジャーナルデータの管理を行っているが、ジャーナルデータの格納ディスクが満杯になってしまうと、バックアップデータを取っていない限り、データを復旧させることはできない。従って、データベースシステムは、ジャーナルデータ用ディスクの容量管理等まで行う必要があり、処理負担が大きくなる。さらに、データの世代管理を行う場合は、複数世代のバックアップデータを作成するため、処理負担が更に増大する。

【０００８】

第２の従来技術では、仮想的なディスク装置にアクセスを切り替えることによ

り、実行中の処理を中断することなく、データ復旧作業を行うことができるが、直前の状態までしか回復させることができず、オペレータが所望する任意の時点にデータを復旧させることができない。

【０００９】

本発明は、上記問題点に鑑みてなされたもので、その目的は、ホストコンピュータ側の処理負担を増大させずに、任意の時点へデータを復旧させることができる外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラムを提供することにある。本発明のさらなる目的は、後述する実施の形態の記載から明らかになるであろう。

【００１０】

【課題を解決するための手段】

上記課題を解決すべく、本発明の第１の観点に従う外部記憶装置は、ホストコンピュータに接続されるもので、ホストコンピュータにより利用されるデータを記憶する記憶手段と、記憶手段を制御する制御手段とを有する。制御手段は、記憶手段に記憶されたデータに関してホストコンピュータにより設定される回復可能時点を登録する登録手段と、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報をホストコンピュータに送信する選択用情報送信手段と、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、を備える。

【００１１】

記憶手段としては、例えば、複数のディスク装置をアレイ状に配置してなる記憶装置を用いることができる。ホストコンピュータは、記憶手段に記憶されるデータに関して、回復可能時点を設定することができる。回復可能時点とは、そのデータを回復させることが可能な時点を示す情報であり、復元ポイントと称することもできる。ホストコンピュータにより定期的に又は不定期に設定される回復可能時点は、登録手段により登録される。

【００１２】

障害が発生等してデータの復旧が必要になった場合、ホストコンピュータは、制御手段に対して、回復可能時点の選択用情報を要求する。この要求に応じて、

選択用情報送信手段は、選択用情報をホストコンピュータに送信する。選択用情報とは、回復可能時点を選択するための情報であり、例えば、一覧形式等で表示可能である。

【００１３】

ホストコンピュータは、受信した選択用情報に基づいて、回復させたい時点を選択する。ホストコンピュータにより選択された回復可能時点は、回復手段に通知される。そして、回復手段は、ホストコンピュータに指定されたデータを指定された時点まで回復させる。回復手段は、例えば、指定された回復時点までのジャーナルデータをバックアップデータに順次反映させていくことにより、データを復旧させることができる。これにより、ホストコンピュータのコンピュータ資源を事実上殆ど使用することなく、外部記憶装置内で任意の時点までデータを回復させることができる。

【００１４】

登録手段は、ホストコンピュータにより設定される任意の複数時点を、回復可能時点として登録可能である。即ち、直前の最新状態のみならず、任意の時点を複数個登録させることができる。例えば、ホストコンピュータは、更新処理（コミット）を要求する度毎に、あるいは、データ操作の区切りがつく度毎に、自動的に又はオペレータによる手動で、回復可能時点を設定することができる。

【００１５】

本発明の一態様では、記憶手段は、ジャーナルデータを取得して記憶するジャーナルデータ記憶手段を有し、登録手段は、ホストコンピュータからの指示に基づいて、ジャーナルデータの所定位置に標識情報を対応付けることにより、回復可能時点を登録するようになっている。即ち、ジャーナルデータは、外部記憶装置内のジャーナルデータ記憶手段が、独自に自動的に収集して記憶する。そして、登録手段は、ホストコンピュータからの設定に基づいて、ジャーナルデータの所定位置に標識情報を対応付けることにより、回復可能時点を登録する。標識情報は、ジャーナルデータ中に含めることもできるし、ジャーナルデータとは別のデータとしてそれぞれ管理し、ユニークな識別コード等で両者を関連付けることもできる。

【００１６】

本発明の一態様では、ジャーナルデータは、少なくとも、書込みデータと、書込み位置と、標識情報としての回復フラグ情報とを含んでなり、登録手段は、ジャーナルデータ中の所定の回復フラグ情報をセットすることにより、回復可能時点を登録するようになっている。

【００１７】

回復フラグを追加してジャーナルデータのデータ構造を拡張する。全てのジャーナルデータには回復フラグをセットするデータ領域が予め含まれている。あるデータについて回復可能時点を設定する場合は、該データに対応する回復フラグをセットする。回復フラグをリセットすれば、設定された回復可能時点を解除することができる。

【００１８】

本発明の一態様では、さらに、記憶手段は、バックアップデータを記憶するバックアップデータ記憶手段を有し、制御手段は、ジャーナルデータ管理手段を有している。そして、ジャーナルデータ管理手段は、ジャーナルデータ記憶手段の空き容量が不足した場合には、ジャーナルデータ記憶手段に記憶されている最古のジャーナルデータをバックアップデータ記憶手段に移し替えて、ジャーナルデータ記憶手段の空き容量を増加させ、かつ、登録された回復可能時点のうち最古の回復可能時点が変更された旨をホストコンピュータに通知する。

【００１９】

データの回復は、例えば、ある時点のバックアップデータに、目標とする時点までのジャーナルデータを順次反映させることにより実現される（ロールフォワード方式）。従って、ジャーナルデータが存在しない場合は、バックアップされた時点にしかデータを戻すことはできない。一方、ジャーナルデータは、データ更新履歴の集合体であり、日々増大する。ジャーナルデータの保存量が、ディスク装置の記憶容量に達すると、それ以上のジャーナルデータを記憶することはできない。そこで、ジャーナルデータの空き容量が不足した場合は、既に蓄積されれているジャーナルデータの中から最も古いデータを必要な量だけバックアップデータに移し替えて、空き容量を確保する。移し替える必要量は、予め設定され

た固定値としても良いし、ジャーナルデータの蓄積速度やバックアップデータ記憶手段の記憶容量等の諸要因に応じて動的に変化させてもよい。ここで、最古のジャーナルデータをバックアップデータに移し替えるとは、最古のジャーナルデータをバックアップデータに反映させた上で、最古のジャーナルデータを削除することを意味する。なお、記憶手段内に未使用の記憶領域がある限り、ジャーナルデータ記憶領域を自動的に拡張させ、未使用の記憶領域が不足した場合に、最古のジャーナルデータをバックアップデータに移し替えるようにしてもよい。

【００２０】

本発明の第２の観点に従う外部記憶装置のデータ回復方法は、ホストコンピュータに接続された外部記憶装置のデータを、該外部記憶装置内で回復させるデータ回復方法であって、記憶されたデータに関してホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録ステップと、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報をホストコンピュータに送信する一覧送信ステップと、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復ステップと、を含んだことを特徴とする。

【００２１】

登録ステップ、一覧送信ステップ、回復ステップは、この順序で実行してもよいし、異なる順序、例えば、並行的に実行してもよい。

【００２２】

本発明の第３の観点に従うプログラムは、ホストコンピュータに接続された外部記憶装置を制御するためのプログラムであって、外部記憶装置は、ホストコンピュータにより利用されるデータを記憶する記憶手段を有し、記憶手段に記憶されたデータに関してホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録手段と、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段とを外部記憶装置のコンピュータ上に実現させる。

【００２３】

本発明の第４の観点に従うプログラムは、外部記憶装置を利用するホストコンピュータを制御するプログラムであって、外部記憶装置に記憶されたデータに関して、任意の複数時点で設定可能な回復可能時点を外部記憶装置に指示し登録させる登録指示手段と、外部記憶装置に登録された回復可能時点の選択用情報を要求する選択用情報要求手段と、外部記憶装置から受信した選択用情報に基づいて、所望のデータを所望の回復可能時点まで回復させるように外部記憶装置に指示する回復指示手段と、をホストコンピュータ上で実現させる。

【００２４】

このプログラムは、例えば、ＡＰＩ（Application Program Interface）のような形で提供可能であり、種々の業務用アプリケーションプログラムから好適に利用されることができる。

【００２５】

本発明に従うプログラムは、例えば、ディスク型記憶媒体、半導体メモリ等の各種記憶媒体に固定して流通に置くこともできるし、あるいは、サーバから通信ネットワークを介して配信することもできる。

【００２６】

【発明の実施の形態】

以下、図１～図１０に基づき、本発明の実施の形態を説明する。

【００２７】

まず最初に、図１に基づいて、外部記憶システムの全体概要を説明する。

【００２８】

先にシステムの全体構成を図１に基づいて説明する。記憶装置システム６０は、記憶デバイス制御装置１０と記憶デバイス３０とを備えて構成されている。記憶デバイス制御装置１０は、情報処理装置２０から受信したコマンドに従って、記憶デバイス３０に対する制御を行う。例えば、記憶デバイス制御装置１０は、情報処理装置２０からデータの入出力要求を受信すると、記憶デバイス３０に記憶されているデータの入出力処理を行う。記憶デバイス３０が備えるディスクドライブにより提供される物理的な記憶領域上には、論理ボリューム（Logical Un

it）（以下、ＬＵと略記）が設定されている。ＬＵは、論理的な記憶領域であり、このＬＵ上にデータは記憶されている。また、記憶デバイス制御装置１０は、情報処理装置２０との間で、記憶装置システム６０を管理するための各種コマンドの授受も行う。

【００２９】

情報処理装置２０は、ＣＰＵ（Central Processing Unit）やメモリ等を備えたコンピュータシステムである。情報処理装置２０のＣＰＵが各種プログラムを実行することにより、種々の機能が実現される。情報処理装置２０は、例えば、パーソナルコンピュータやワークステーションである場合もあるし、メインフレームコンピュータの場合もある。図１では、説明の便宜上、第１～第５の５台の情報処理装置を図示する。各情報処理装置２０を識別するために、図１中では「情報処理装置１」、「情報処理装置２」等のように連番を付し、第１～第５の情報処理装置２０とする。後述のチャネル制御部１１及びディスク制御部１４も同様に連番を付して区別する。

【００３０】

第１～第３の情報処理装置２０は、ＬＡＮ（Local Area Network）４０を介して、記憶デバイス制御装置１０と接続されている。ＬＡＮ４０は、例えば、インターネットとすることもできるし、専用のネットワークとすることもできる。第１～第３の情報処理装置２０と記憶デバイス制御装置１０との間のデータ通信は、ＬＡＮ４０を介して、例えば、TCP/IP（Transmission Control Protocol/Internet Protocol）プロトコルに従って行われる。第１～第３の情報処理装置２０からは、記憶装置システム６０に対して、ファイル名指定によるデータアクセス要求（ファイル単位でのデータ入出力要求である。以下、「ファイルアクセス要求」と略記）が送信される。

【００３１】

ＬＡＮ４０には、バックアップデバイス７１が接続されている。バックアップデバイス９１としては、例えば、ＭＯ（magneto-optic：光磁気型記憶装置）、ＣＤ－Ｒ（CD-Recordable：読み書き可能なコンパクトディスク）、DVD-RAM（Digital Versatile Disk-RAM：読み書き可能なDVD）等のディスク系記憶デバイス

や、例えば、ＤＡＴ（Digital Audio Tape）テープ、カセットテープ、オープンテープ、カートリッジテープ等のテープ系記憶デバイスを用いることができる。バックアップデバイス７１は、ＬＡＮ４０を介して記憶デバイス制御装置１０との間で通信を行うことにより、記憶デバイス３０に記憶されているデータのバックアップデータを記憶する。また、バックアップデバイス７１は、第１の情報処理装置２０と接続されるように構成することもできる。この場合は、第１の情報処理装置２０を介して、記憶デバイス３０に記憶されているデータのバックアップデータを取得するようにする。

【００３２】

記憶デバイス制御装置１０は、第１～第４のチャネル制御部１１により、ＬＡＮ４０を介して第１～第３の情報処理装置２０やバックアップデバイス７１との間で通信を行う。第１～第４のチャネル制御部１１は、第１～第３の情報処理装置２０からのファイルアクセス要求を個々に受け付ける。即ち、第１～第４のチャネル制御部１１には、それぞれＬＡＮ４０上のネットワークアドレス（例えば、ＩＰアドレス）が割り当てられており、第１～第４の各チャネル制御部１１はそれぞれが個別にＮＡＳ（Network Attached Storage）として振る舞うようになっている。第１～第４のチャネル制御部１１は、それぞれが独立したＮＡＳであるかのように、第１～第３の情報処理装置２０に対しＮＡＳとしてのサービスを提供可能である。以下、第１～第４のチャネル制御部１１をＣＨＮと略す場合がある。このように、１台の記憶装置システム６０内にそれぞれ個別にＮＡＳとしてのサービスを提供する第１～第４のチャネル制御部１１を備えるように構成したことにより、従来、独立したコンピュータで個々に運用されていたＮＡＳサーバが１台の記憶装置システム６０に集約される。そして、これにより、記憶装置システム６０の統括的な運用が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化を図ることができる。

【００３３】

なお、記憶デバイス制御装置１０の第１～第４のチャネル制御部１１は、例えば、一体的にユニット化された回路基板上に形成されたハードウェア、このハードウェアにより実行されるＯＳ（Operating System）、このＯＳ上で動作するア

プリケーションプログラム等のソフトウェアにより実現される。記憶装置システム６０では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。従って、記憶装置システム６０を用いることにより、柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズにきめ細やかに対応可能となる。

【００３４】

第３及び第４の情報処理装置２０は、ＳＡＮ（Storage Area Network）５０を介して、記憶デバイス制御装置１０と接続されている。ＳＡＮ５０は、記憶デバイス３０が提供する記憶領域におけるデータの管理単位であるブロックを単位として、第３及び第４の情報処理装置２０との間でデータの授受を行うためのネットワークである。ＳＡＮ５０を介して行われる第３及び第４の情報処理装置２０と記憶デバイス制御部１０との間の通信は、一般にファイバチャネルプロトコルに従う。第３及び第４の情報処理装置２０からは、記憶装置システム６０に対して、ファイバチャネルプロトコルに従ってブロック単位でのデータアクセス要求（以下、ブロックアクセス要求と略記）が送信される。

【００３５】

ＳＡＮ５０には、ＳＡＮ対応のバックアップデバイス７０が接続されている。ＳＡＮ対応バックアップデバイス７０は、ＳＡＮ５０を介して記憶デバイス制御装置１０との間で通信を行うことにより、記憶デバイス３０に記憶されているデータのバックアップデータを記憶する。

【００３６】

記憶デバイス制御装置１０は、第５及び第６のチャネル制御部１１により、ＳＡＮ５０を介して第３及び第４の情報処理装置２０及びＳＡＮ対応バックアップデバイス７０との間の通信を行う。以下、第５及び第６のチャネル制御部１１をＣＨＦと略記する場合がある。

【００３７】

また、第５の情報処理装置２０は、ＬＡＮ４０やＳＡＮ５０等のネットワークを介さずに、記憶デバイス制御装置１０と直接的に接続されている。第５の情報処理装置２０としては、例えば、メインフレームコンピュータとすることができ

るが、もちろんこれに限定されない。第５の情報処理装置２０と記憶デバイス制御装置１０との間の通信は、例えば、FICON（Fibre Connection）（登録商標）やESCON（Enterprise System Connection）（登録商標）、ACONARC（Advanced Connection Architecture）（登録商標）、FIBARC（Fibre Connection Architecture）（登録商標）等の通信プロトコルに従う。第５の情報処理装置２０からは、記憶装置システム６０に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

【００３８】

記憶デバイス制御装置１０は、第７及び第８のチャネル制御部１１により、第５の情報処理装置２０との間で通信を行う。以下、第７及び第８のチャネル制御部１１をＣＨＡと略す場合がある。

【００３９】

ＳＡＮ５０には、記憶装置システム６０の設置場所（プライマリサイト）から遠隔した場所（セカンダリサイト）に設置される他の記憶装置システム６１が接続されている。他の記憶装置システム６１は、レプリケーションまたはリモートコピーの機能におけるデータ複製先の装置として利用される。なお、他の記憶装置システム６１は、ＳＡＮ５０以外にも例えばＡＴＭ（Asynchronous Transfer Mode）等の通信回線を介して記憶装置システム６０に接続される場合もある。この場合には、上記通信回線を利用するためのインターフェース（チャネルエクステンダ）を備えたチャネル制御部１１が採用される。

【００４０】

次に、記憶デバイス３０の構成について説明する。記憶デバイス３０は、多数のディスクドライブ（物理ディスク）を備えており、情報処理装置２０に対して記憶領域を提供する。データは、論理的記憶領域であるＬＵに記憶されている。ディスクドライブとしては、例えば、ハードディスク装置、フレキシブルディスク装置、半導体記憶装置等の種々のデバイスを用いることができる。なお、記憶デバイス３０は、例えば、複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置２０に対しては、ＲＡＩＤ（Redundant Array of Independent（Inexpensive）Disks）により管理された

複数のディスクドライブにより記憶領域を提供することができる。

【００４１】

記憶デバイス制御装置１０と記憶デバイス３０とは、図１に示すように、直接的に接続してもよいし、ネットワークを介して間接的に接続するようにしてもよい。さらに、記憶デバイス３０は、記憶デバイス制御装置１０と一体のものとして構成することもできる。

【００４２】

記憶デバイス３０に設定されるＬＵには、情報処理装置２０からアクセス可能なユーザＬＵや、チャネル制御部１１の制御のために使用されるシステムＬＵ等がある。システムＬＵには、ＣＨＮ１１で実行されるＯＳも格納される。また、各ＬＵには、各チャネル制御部１１が予め対応付けられている。これにより、各チャネル制御部１１毎にアクセス可能なＬＵがそれぞれ割り当てられている。また、上記対応付けは、複数のチャネル制御部１１で一つのＬＵを共有するように設定することもできる。なお、以下の説明において、ユーザＬＵをユーザディスク、システムＬＵをシステムディスクと記す場合がある。また、複数のチャネル制御部１１により共有されるＬＵを、共有ＬＵまたは共有ディスクと記す場合がある。

【００４３】

次に、記憶デバイス制御装置１０の構成を説明する。記憶デバイス制御装置１０は、チャネル制御部１１、共有メモリ１２、キャッシュメモリ１３、ディスク制御部１４、接続部１５及び管理端末１６を備えている。

【００４４】

チャネル制御部１１は、情報処理装置２０との間で通信を行うための通信インターフェースを有し、情報処理装置２０との間でデータ入出力コマンド等を授受する機能を備えている。例えば、ＣＨＮ１１は、第１～第３の情報処理装置２０からのファイルアクセス要求を受け付ける。これにより、記憶装置システム６０は、ＮＡＳとしてのサービスを第１～第３の情報処理装置２０に提供することができる。また、ＣＨＦ１１は、第３及び第４の情報処理装置２０からのファイバチャネルプロトコルに従ったブロックアクセス要求を受け付ける。これにより、

記憶装置システム６０は、高速アクセス可能なデータ記憶サービスを第３及び第４の情報処理装置２０に対して提供することができる。また、ＣＨＡ１１は、第５の情報処理装置２０からのFICONやESCON、ACONARC、FIBARC等のプロトコルに従ったブロックアクセス要求を受け付ける。これにより、記憶装置システム６０は、第５の情報処理装置２０のようなメインフレームコンピュータ等に対してもデータ記憶サービスを提供することができる。

【００４５】

各チャネル制御部１１は、管理端末１６と共に内部ＬＡＮ１７で接続されている。これにより、チャネル制御部１１に実行させるプログラム等を、管理端末１６からチャネル制御部１１に送信してインストールさせることも可能となっている。チャネル制御部１１の構成については、さらに後述する。

【００４６】

接続部１５は、各チャネル制御部１１、共有メモリ１２、キャッシュメモリ１３、各ディスク制御部１４を相互に接続する。チャネル制御部１１、共有メモリ１２、キャッシュメモリ１３及びディスク制御部１４間でのデータやコマンドの授受は、接続部１５を介することにより行われる。接続部１５は、例えば、高速スイッチングによりデータ伝送を行う超高速クロスバスイッチ等の高速バスで構成される。チャネル制御部１１同士を高速バスで接続することにより、個々のコンピュータ上で動作するＮＡＳサーバをＬＡＮを介して接続する場合よりも、チャネル制御部１１間の通信パフォーマンスが向上する。また、これにより、高速なファイル共有機能や高速フェイルオーバ等が可能となる。

【００４７】

共有メモリ１２及びキャッシュメモリ１３は、各チャネル制御部１１及び各ディスク制御部１４により共有される記憶メモリである。共有メモリ１２は、主に制御情報やコマンド等を記憶するために利用される。キャッシュメモリ１３は、主にデータを記憶するために利用される。

【００４８】

例えば、あるチャネル制御部１１が情報処理装置２０から受信したデータ入出力コマンドが書込みコマンドであった場合、当該チャネル制御部１１は、書込み

コマンドを共有メモリ１２に書き込むと共に、情報処理装置２０から受信した書込みデータをキャッシュメモリ１３に書き込む。一方、ディスク制御部１４は、共有メモリ１２を監視している。ディスク制御部１４は、共有メモリ１２に書込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ１３から書込みデータを読出し、読み出したデータを記憶デバイス３０に書き込む。

【００４９】

ディスク制御部１４は、記憶デバイス３０の制御を行う。例えば、上述のように、ディスク制御部１４は、チャネル制御部１１が情報処理装置２０から受信した書込みコマンドに従って、記憶デバイス３０へデータの書込みを行う。また、ディスク制御部１４は、チャネル制御部１１から送信された論理アドレス指定によるＬＵへのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。ディスク制御部１４は、記憶デバイス３０における物理ディスクがＲＡＩＤにより管理されている場合は、ＲＡＩＤ構成に従ったデータのアクセスを行う。また、ディスク制御部１４は、記憶デバイス３０に記憶されたデータの複製管理の制御及びバックアップ制御も行う。さらに、ディスク制御部１４は、災害発生時のデータ消失防止（ディザスタリカバリ）等を目的として、プライマリサイトの記憶装置システム６０のデータの複製をセカンダリサイトに設置された他の記憶装置システム６１にも記憶させる制御（レプリケーション機能またはリモートコピー機能と呼ばれる）等も行う。

【００５０】

各ディスク制御部１４は、管理端末１６と共に内部ＬＡＮ１７を介して接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部１４に実行させるプログラム等を、管理端末１６からディスク制御部１４に送信してインストールさせることが可能となっている。

【００５１】

次に、管理端末１６について説明する。管理端末１６は、記憶装置システム６０を保守・管理するためのコンピュータである。管理端末１６を操作することにより、例えば、記憶デバイス３０内の物理ディスク構成の設定や、ＬＵの設定、

チャネル制御部１１で実行させるためのプログラムのインストール等を行うことができる。ここで、記憶デバイス３０内の物理ディスク構成の設定としては、例えば、物理ディスクの増設や減設、ＲＡＩＤ構成の変更（ＲＡＩＤ１からＲＡＩＤ５への変更等）などを挙げることができる。さらに、管理端末１６からは、記憶装置システム６０の動作状態の確認や故障部位の特定、チャネル制御部１１で実行されるＯＳのインストール等の作業を行うこともできる。また、管理端末１６は、ＬＡＮや電話回線等で外部保守センタと接続されており、外部保守センタから管理端末１６を利用して記憶装置システム６０の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は、例えば、ＯＳやアプリケーションプログラム、ドライバソフトウェア等から通知される。この通知は、例えば、ＨＴＴＰ（HyperText Transfer Protocol）プロトコルやＳＮＭＰ（Simple Network Management Protocol）プロトコル、電子メール等により行うことができる。これらの設定や制御は、管理端末１６上で動作するウェブサーバが提供するウェブページをユーザインターフェースとして、オペレータ等が操作することにより行うことができる。オペレータ等は、管理端末１６を操作して、障害監視の対象や内容を設定したり、障害通知先を設定等する。

【００５２】

管理端末１６は、記憶デバイス制御装置１０内に内蔵させる構成でもよいし、記憶デバイス制御装置１０に外付けする構成でもよい。また、管理端末１６は、記憶デバイス制御装置１０及び記憶デバイス３０の保守・管理を専用に行うコンピュータとして構成してもよいし、あるいは、汎用コンピュータに保守・管理機能を持たせることにより構成してもよい。

【００５３】

次に、図２を参照して本発明によるデータ回復方法の一例を説明する。図２は、図１と共に述べた記憶装置システムの要部を抜き出した概略構成図である。図２に示す外部記憶システムは、それぞれ後述するように、ホストコンピュータ１０と外部記憶装置とに大別され、外部記憶装置は、ディスク制御装置２００と大容量記憶装置４００とに大別される。ここで、図１と図２との対応関係を簡単に説明すると、図１中の記憶装置システム６０が図２中のディスク制御装置２００

に、図１中のチャネル制御部１１が図２中のチャネルポート２１０及びマイクロプロセッサ２２０に、図１中の共有メモリ１２及びキャッシュメモリ１３が図２中のバッファメモリ２３０に、図１中の接続部１５がバスやスイッチ類等（図示せず）に、図１中のディスク制御部１４が図２中のマイクロプロセッサ２２０に、図１中の記憶デバイス３０が図２中の記憶装置４００に、図１中の情報処理装置２０が図２中のホストコンピュータ１００に、それぞれ対応する。マイクロプロセッサ２２０は、チャネル制御部１１またはディスク制御部１４のいずれの側に存在してもよい。

【００５４】

ホストコンピュータ１００は、例えば、パーソナルコンピュータやワークステーション等から構成されるもので、データベースを扱うアプリケーションプログラム１１０（以下、アプリケーションと略記）を有する。また、図示を省略しているが、ホストコンピュータ１００は、例えば、ポインティングデバイス、キーボードスイッチ、モニタディスプレイ等を通じてオペレータと情報を交換するためのユーザインターフェースを備えている。アプリケーション１１０は、ディスク制御装置２００を介して記憶装置４００内のデータにアクセスすることにより、所定の業務を処理する。

【００５５】

ディスク制御装置２００は、記憶装置４００を制御するもので、チャネルポート２１０、マイクロプロセッサ２２０及びバッファメモリ２３０を備えている。

【００５６】

マイクロプロセッサ２２０は、チャネルポート２１０を介して、ホストコンピュータ１００と双方向のデータ通信を行う。マイクロプロセッサ２２０は、ディスク制御プログラム３００を実行する。ディスク制御プログラム３００には、書き込み制御処理３１０、書込みデータ処理３２０、ディスク管理処理３３０、データ回復制御処理３４０、データ回復処理３５０、データ同期処理３６０が含まれている。

【００５７】

主要な処理については、詳細をさらに後述するが、書込み制御処理３１０は、

主としてデータ書込み時の書込み制御情報（ジャーナル制御情報）を管理するものである。書込みデータ処理３２０は、所定のディスク装置へのデータ書込みを行うものである。ディスク管理処理３３０は、主としてジャーナルデータ格納ディスク４３０の管理を行うものである。データ回復制御処理３４０は、ホストコンピュータ１００から設定される回復契機の登録と、登録された回復契機のリストデータをホストコンピュータ１００に送信するものである。データ回復処理３５０は、指定されたディスク装置のデータを指定された時点まで回復させるものである。データ同期処理３６０は、ホストコンピュータ１００からの指示に応じて、データのバックアップ処理を行うものである。

【００５８】

バッファメモリ２３０には、例えば、回復データ情報Ｄ１０、ジャーナルデータＤ２０、書込み制御情報Ｄ３０、更新データＤ４０が記憶されている。回復データ情報Ｄ１０は、データの回復処理（復旧処理）の履歴情報であり、例えば、データ回復先や回復時点等を記録している。ジャーナルデータＤ２０は、データ操作の更新履歴であり、バッファメモリ２３０から順次ジャーナル格納ディスク４３０に移される。書込み制御情報Ｄ３０は、任意の時点にデータを回復させるために必要な情報を含んでいる。更新データＤ４０は、アプリケーション１１０により更新が指示されたデータであり、バッファメモリ２３０からデータ格納ディスク４１０に移される。なお、以上のデータは、バッファメモリ２３０上に同時に存在する必要はない。また、説明の便宜上、バッファメモリ２３０を単一のメモリのように示したが、例えば、複数種類のメモリ装置の集合体として構成してもよい。

【００５９】

大容量記憶装置４００は、データ格納ディスク４１０、バックアップデータ格納ディスク４２０及びジャーナルデータ格納ディスク４３０を備えている。データ格納ディスク４１０には、現在使用中の最新データ（実データ）が格納されている。バックアップデータ格納ディスク４２０には、ある時点のバックアップデータが格納されている。ジャーナルデータ格納ディスク４３０には、ジャーナルデータが格納されている。なお、各ディスク４１０～４３０は、正確にはディス

ク装置であり、それぞれ複数のディスクを備えている。以下、データ格納ディスクをデータディスク、バックアップデータ格納ディスクをバックアップディスク、ジャーナルデータ格納ディスクをジャーナルディスクと呼ぶ。

【００６０】

図３は、ジャーナルデータＤ２０及び書込み制御情報Ｄ３０の概略構造を示すデータ構造図である。

【００６１】

本実施の形態によるジャーナルデータＤ２０は、書込み制御情報Ｄ３０と更新データ（書込みデータ）Ｄ４０とが含まれている。書込み制御情報Ｄ３０は、ジャーナル制御情報としての機能を果たすもので、例えば、データ書込み位置Ｄ３１、データサイズＤ３２、タイムスタンプＤ３３、回復フラグＤ３４、その他制御情報Ｄ３５等の情報を含んでいる。データ書込み位置Ｄ３１は、どのディスク装置のどこにデータが書き込まれたかを示す位置情報である。データサイズＤ３２は、書き込まれたデータのサイズを示す情報である。タイムスタンプＤ３３は、データ書込み時刻を示す情報である。回復フラグＤ３４は、回復可能な時点（復元ポイント）であることを示す標識情報であり、回復フラグＤ３４をセットすると、回復可能なデータとして設定され、回復フラグＤ３４をリセットすると、復元ポイントの設定が解除される。その他の制御情報Ｄ３５には、例えば、書込み制御情報Ｄ３０を一意に特定するための制御番号やデータ種別等のその他必要な情報が含まれる。

【００６２】

本実施形態では、図３に示すように、ジャーナルデータＤ２０の構造を独自に拡張し、ジャーナルデータＤ２０内に回復フラグＤ３４を設けている。これにより、少量のデータを追加するだけで任意の時点を回復可能時点として自由に設定することができ、任意の時点にデータを回復させることができる。但し、これに限らず、ジャーナルデータＤ２０と回復フラグＤ３４とを分離し、ユニークなＩＤ（識別コード）等で両者を対応付ける構成でもよい。

【００６３】

次に、図４は、ホストコンピュータ１００及びディスク制御装置２００のプロ

グラム構造の概略を示すブロック図である。

【００６４】

アプリケーション１１０は、ホストコンピュータ１００のＯＳ１２０を介してディスク制御プログラム３００と双方向のデータ通信を行う。ＯＳ１２０は、ＡＰＩ（Application Program Interface）群１３０を有する。ＡＰＩ群１３０には、データ書込み用ＡＰＩ１３１、回復契機通知用ＡＰＩ１３２、回復契機リスト取得要求用ＡＰＩ１３３、回復指示用ＡＰＩ１３４が含まれている。アプリケーション１１０は、これらＡＰＩ１３１～１３４を適宜呼び出して利用することにより、所望の時点を回復契機として設定し、設定済の回復契機リストを読み出し、所望の時点を選択してデータの回復を指示することができる。

【００６５】

図４を参照しつつ全体の動作を簡単に説明する。アプリケーション１１０が、データ書込み用ＡＰＩ１３１を介して、ディスク制御装置２００にデータ更新要求（コミット要求）を指示すると（Ｓ１）、ディスク制御プログラム３００の書込み制御処理３１０は、書込みデータ処理３２０を介して所定のディスクにデータを書き込ませ、更新要求を処理した旨をアプリケーション１１０に通知する（Ｓ２）。

【００６６】

アプリケーション１１０は、業務処理中に、例えば、定期的に又は不定期に所望の時点を、回復可能な時点としての回復契機（復元ポイント）として設定することができる。アプリケーション１１０は、回復契機通知用ＡＰＩ１３２を呼び出すことにより、回復契機を設定するデータをディスク制御装置２００に対し指示する（Ｓ３）。回復契機が通知されると、ディスク制御プログラム３００のデータ回復制御処理３４０は、指定されたデータの回復フラグをセットし、回復契機が設定された旨をアプリケーション１１０に通知する（Ｓ４）。

【００６７】

障害発生等の要因によりデータを回復させる場合は、アプリケーション１１０は、回復契機リスト取得要求用ＡＰＩ１３３を呼び出し、回復可能な時点のリスト情報をディスク制御装置２００に要求する（Ｓ５）。リストを要求されると、

データ回復制御処理３４０は、ジャーナルディスク４３０を検査して回復フラグがセットされたデータの情報を取得し、回復契機リストを作成する。データ回復制御処理３４０は、回復契機リストをアプリケーション１１０に返信する（Ｓ６）。

【００６８】

アプリケーション１１０は、メモリ１４０に格納された回復契機リストを参照し、回復を希望する時点を少なくとも１つ選択する。アプリケーション１１０は、回復指示用ＡＰＩ１３４を呼び出すことにより、希望する時点まで所定ディスクのデータを回復させるように、ディスク制御装置２００に指示する（Ｓ８）。データ回復処理３５０は、アプリケーション１１０から回復指示を受けると、バックアップディスク４２０及びジャーナルディスク４３０を用いて、指定されたデータを指定された時点まで回復させる。回復処理３５０は、回復処理が完了した旨をアプリケーション１１０に通知する（Ｓ９）。

【００６９】

次に、図５～図９を参照して各部の詳細な制御を説明する。まず、図５は、書込み制御処理を示すフローチャートである。なお、以下の説明でも同様であるが、図示するフローチャートは発明の理解のために動作の要部を示すものであり、実際のプログラムとは相違する可能性がある。図中、「ステップ」を「Ｓ」と略記する。

【００７０】

アプリケーション１１０が書込み要求を出すと、バッファメモリ２３０上のデータＤ４０が更新されると共に（Ｓ２１）、バッファメモリ２３０上の書込み制御情報Ｄ３０が更新される（Ｓ２２）。次に、ジャーナルディスク４３０に十分な空き容量があるか否かを判定する（Ｓ２３）。例えば、ジャーナルディスク４３０の現在の空き容量が、これから書き込もうとするデータのデータサイズを上回っているか否かで判定することができる。ジャーナルディスク４３０の空き容量が不足している場合は（S23:NO）、図６と共に後述するジャーナルディスク管理処理を実行して空き容量を確保し（Ｓ２４）、必要な場合は、バッファメモリ２３０上の書込み制御情報を更新する（Ｓ２５）。必要な場合とは、例えば、後

述のジャーナル自動拡張により、ジャーナルデータの書込み位置が変動等した場合である。

【００７１】

ジャーナルディスク４３０に十分な空き容量が存在した場合（S23:YES）及びジャーナルディスク４３０に十分な空き容量が確保された場合は、書込みデータＤ４０及び書き込み制御情報Ｄ３０（即ち、ジャーナルデータＤ２０）をジャーナルディスク４３０に追加して書き込む（Ｓ２６）。また、バッファメモリ２３０上の書込みデータＤ４０をデータディスク４１０の所定位置に書き込み（Ｓ２７）、データ書込みが完了した旨をホストコンピュータ１００（正確にはホストコンピュータ１００上のアプリケーション１１０である。以下同様）に通知する（Ｓ２８）。

【００７２】

なお、Ｓ２６及びＳ２７は、本書込み制御処理とは別契機（非同期）に行っても良い。その場合、例えば、バッファメモリ上の該データにディスクへ反映したか否かのフラグを設けることにより管理することができる。

【００７３】

そして、バックアップ更新フラグがオンになっているか否かを判定する（Ｓ２９）。バックアップ更新フラグとは、ジャーナルディスク４３０の空き容量を確保するために、最古のジャーナルデータをバックアップディスク４２０に移し替えたことを示す標識情報である。ジャーナルデータの移し替えにより、バックアップデータから回復可能な最古の時点が変更されるため、バックアップ更新フラグがオン状態にセットされている場合は（S29:YES）、バックアップデータが更新された旨をホストコンピュータ１００に通知する（Ｓ３０）。バックアップ更新をホストコンピュータ１００に通知した後、バックアップ更新フラグをオフ状態にリセットさせる（Ｓ３１）。

【００７４】

次に、図６は、図５中のジャーナルディスク管理処理Ｓ２４の詳細を示すフローチャートである。まず、ジャーナルディスク４３０の自動拡張モードが設定されているか否かを判定する（Ｓ４１）。自動拡張モードとは、未使用のディスク

、未使用の記憶領域を探索してジャーナルディスク４３０の論理的サイズを自動的に拡大させるモードである。

【００７５】

自動拡張モードが設定されていない場合は（S41:NO）、ジャーナルディスク４３０に記憶されているジャーナルデータのうち最も古いデータを選択して、バックアップディスク４２０に反映させる（Ｓ４２）。バックアップディスク４２０に移し替えられた最古のジャーナルデータは、ジャーナルディスク４３０から消去される（Ｓ４３）。これによりジャーナルディスク４３０の空き容量が増加する。ジャーナルディスク４３０の空き容量が所定値に達するまで、最古のジャーナルデータから順番にバックアップディスク４２０に移し替える（Ｓ４４）。ジャーナルディスク４３０の空き容量が所定値に達した場合は（S44:YES）、バックアップ更新フラグをオン状態にセットする（Ｓ４５）。これにより、図５中のＳ３０に示したように、バックアップデータが更新され、バックアップデータから回復可能な最古の時点が変更された旨がホストコンピュータ１００に通知される。なお、Ｓ４４中の所定値は、予め設定された固定値であっても良いし、例えば、バックアップディスクの空き容量やデータディスク４１０に書き込まれるデータサイズ等に応じて動的に変更される値であってもよい。

【００７６】

一方、ジャーナルディスク４３０の自動拡張モードが設定されている場合は（S41:YES）、接続されているディスク装置の中から未使用の記憶エリア（未使用エリアと呼ぶ）を検索し、ジャーナルデータを保存可能な未使用エリアが存在するか否かを判断する（Ｓ４６，Ｓ４７）。未使用エリアが発見されなかった場合は（S47:NO）、Ｓ４２に移り、上述のように最古のジャーナルデータをバックアップディスク４２０に移し替えることにより、ジャーナルディスク４３０に空き容量を確保する。未使用エリアが発見された場合は（S47:YES）、発見された未使用エリアをジャーナルディスクとして利用すべく、ジャーナルディスク４３０の論理サイズを拡張すると共に、ディスク管理マップを更新する（Ｓ４８）。そして、ジャーナルディスク４３０の論理サイズ拡張により生じた空き容量が所定値に達したか否かを判定し（Ｓ４９）、ジャーナルディスク４３０の空き容量が

所定値に達するまで、Ｓ４６～Ｓ４９の処理を繰り返しながら、未使用エリアをジャーナルデータの記憶エリアとして自動的に拡張させる。

【００７７】

次に、図７は、ホストコンピュータ１００から指示される回復契機の登録処理を示す。上述の通り、本実施形態では、ホストコンピュータ１００は、任意の時点を回復可能な契機（復元ポイント）として複数個設定することができる。

【００７８】

登録すべき回復契機がホストコンピュータ１００からディスク制御装置２００に通知されると、データ回復制御処理３４０は、ジャーナルディスク４３０に記憶されている最新データの位置を検索し（Ｓ５１）、最新の書込みデータに対応する書込み制御情報中の回復フラグをオン状態にセットして更新する（Ｓ５２）。そして、ホストコンピュータ１００に、回復契機の設定が完了した旨を報告すると共に、書込み制御情報を特定するための制御番号を通知する（Ｓ５３）。このように、ホストコンピュータ１００のアプリケーション１１０は、データ書込み時に、任意の時点のデータについて回復契機を設定指示できるようになっている。

【００７９】

次に、図８は、ホストコンピュータ１００からの要求に応じて、回復契機のリスト情報を返信する回復契機リストの送信処理を示す。まず、ジャーナルディスク４３０のうち、ホストコンピュータ１００から回復を指定されたデータに対応するディスクを選択し、選択されたディスク中の最古のジャーナルデータにポインタを合わせる（Ｓ６１）。

【００８０】

そして、最古のジャーナルデータから読み込み（Ｓ６２）、読み込んだジャーナルデータに関する書き込み制御情報中の回復フラグがオン状態にセットされているか否かを検査し（Ｓ６３）、回復フラグがセットされている場合は、読み込んだジャーナルデータを回復契機のリスト情報に追加して記録する（Ｓ６４）。Ｓ６１で選択されたディスクに記憶されている最終データを読み出すまで、上記Ｓ６２～Ｓ６４が繰り返される（Ｓ６５）。このようにして、指定されたデータ

に対応するジャーナルデータを最古のデータから最新のデータまで順番に検査し、回復フラグがセットされているジャーナルデータを抽出して回復契機リストを生成する。生成された回復契機リストは、完了報告と共にまたは非同期にホストコンピュータ１００に送信される（Ｓ６６）。

【００８１】

次に、図９は、データ回復処理を示す。ホストコンピュータ１００上のアプリケーションプログラム１１０は、図８に示す処理により取得した回復契機のリスト情報に基づいて、所望の時点までデータの回復を指示することができる。

【００８２】

ホストコンピュータ１００から回復指示が通知されると、データ回復処理３５０は、バックアップディスク４２０及びジャーナルディスク４３０のうち、回復が指定されたデータに対応するディスクをそれぞれ選択する（Ｓ７１）。

【００８３】

次に、ホストコンピュータ１００からデータ回復先として指定されたディスクがバックアップディスク４２０であるか否かを判定する（Ｓ７２）。つまり、本実施形態では、バックアップディスク４２０以外の他のディスク装置に、指定された時点までのデータを回復させることができるようになっている。回復先として指定されたディスク装置がバックアップディスク４２０以外の他のディスク装置である場合は、バックアップディスク４２０に記憶されているバックアップデータを、指定されたディスク装置にコピーし、データ回復の土台となるバックアップデータの準備を完了させる（Ｓ７３）。

【００８４】

次に、ジャーナルディスク４３０から最古のジャーナルデータを検索し（Ｓ７４）、最古のジャーナルデータから順番にデータを読み出して、回復先に指定されたディスクの記憶内容に反映させていく（Ｓ７５）。ホストコンピュータ１００から指定された時点までデータが回復されるまで、ジャーナルデータを読み出して、回復先ディスクの記憶内容を更新させる（Ｓ７６）。

【００８５】

指定された時点までデータが回復した場合は、ホストコンピュータ１００にデ

ータ回復が完了した旨を通知する（Ｓ７７）。また、回復時点及び回復先等の情報を回復データ情報Ｄ１０に記録する（Ｓ７８）。

【００８６】

本実施の形態によれば、外部記憶装置内で自動的にデータの回復を行うため、ホストコンピュータ１００のコンピュータ資源をデータ回復処理のために消費することがなく、ホストコンピュータ１００上の他の業務処理の効率を低下させることがない。特に、大容量の外部記憶装置を用いるアプリケーション１１０では、大規模なデータを取り扱うため、データ回復処理の負担が大きくなり、コンピュータ資源を多量に消費する。従って、ホストコンピュータ１００上で行われる他の業務の処理速度が低下する上に、データ回復完了までの処理時間も長くなる。しかし、本実施形態では、回復契機の設定指示、回復契機リストの取得要求及び回復指示という僅かな処理だけをホストコンピュータ１００で実行し、実際のデータ回復処理を外部記憶装置に委ねる構成のため、ホストコンピュータ１００の負担を軽減することができる。外部記憶装置内でデータの回復を行っている間、ホストコンピュータ１００は、他の業務を効率的に処理することができる。

【００８７】

また、任意の複数時点を回復契機として設定可能であり、所望の時点までデータを回復可能なため、単純に直前のデータに回復させるだけの従来技術とは異なり、利便性が高い。

【００８８】

さらに、本実施形態では、回復契機の設定指示や回復契機リストの取得要求等をホストコンピュータ１００側から行うためのＡＰＩ１３１～１３４を用意したので、これら独自のＡＰＩをホストコンピュータが備えるだけで、本発明に従う外部記憶装置を利用可能となる。

【００８９】

また、本実施形態では、外部記憶装置内でジャーナルデータを自動的に収集すると共に、ジャーナルディスク４３０の空き容量管理も行うため、ジャーナルディスク４３０が満杯となってデータ回復が不能となるのを未然に防止できる。

【００９０】

また、本実施形態では、ジャーナルデータＤ２０のデータ構造を拡張し、ジャーナルデータＤ２０内に（ジャーナル制御情報としての書込み制御情報Ｄ３０内に）回復フラグを設定する構成のため、比較的簡易な構成でありながら、任意の複数時点へのデータ回復を実現することができる。

【００９１】

図１０は、本発明の第２の実施形態を示す。本実施形態では、複数世代のデータ管理を行っている。即ち、最新のデータを保持するデータディスク４１０に加えて、１世代前のデータを格納する１世代前データディスク４１０（１ＧＡ）、２世代前のデータを格納する２世代前データディスク４１０（２ＧＡ）等のように、複数世代でデータを管理できる。

【００９２】

例えば、１世代前データディスク４１０（１ＧＡ）にバックアップディスク４２０の記録内容をリストアした後、ジャーナルディスク４３０に記憶されているデータｄＢのジャーナルデータを読み出して１世代前データディスク４１０（１ＧＡ）に反映させれば、１世代前のデータに戻すことができる。同様に、２世代前データディスク４１０（２ＧＡ）に、バックアップデータをコピーしてからデータｄＢ及びデータｄＣのジャーナルデータを反映させることにより、２世代前のデータに戻すことができる。このように、複数世代でデータを管理する場合も、本発明に従えば、ホストコンピュータ１００に処理負担をかけることなく、外部記憶装置内で複数世代のデータを構築し管理することができる。

【００９３】

なお、本発明は、上述した各実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。

【図面の簡単な説明】

【図１】

本発明の第１の実施の形態に係る外部記憶システムの概略構成図である。

【図２】

図１に示す記憶装置システムの概略を示すブロック図である。

【図３】

ジャーナルデータ及び書込み制御情報の構造を示すデータ構造図である。

【図４】

ホストコンピュータ及びディスク制御装置のプログラム構造を示すブロック図である。

【図５】

書込み制御処理を示すフローチャートである。

【図６】

ジャーナルディスク管理処理を示すフローチャートである。

【図７】

ホストコンピュータから回復契機を通知された場合のデータ回復制御処理を示すフローチャートである。

【図８】

ホストコンピュータから回復契機リストの送信を要求された場合のデータ回復制御処理を示すフローチャートである。

【図９】

ホストコンピュータから回復を指示された場合のデータ回復処理を示すフローチャートである。

【図１０】

複数世代でデータ管理を行う場合の模式図である。

【符号の説明】

１０　記憶デバイス制御装置

１１　チャネル制御部

１２　共有メモリ

１３　キャッシュメモリ

１４　ディスク制御部

１５　接続部

１６　管理端末

１７　内部ＬＡＮ

２０　情報処理装置

３０　記憶デバイス

４０　ＬＡＮ

５０　ＳＡＮ

６０，６１　記憶装置システム

７０，７１　バックアップデバイス

１００　ホストコンピュータ

１１０　アプリケーションプログラム

１２０　ＯＳ

１３０　ＡＰＩ

１３１　データ書込み用ＡＰＩ

１３２　回復契機通知用ＡＰＩ

１３３　回復契機取得要求用ＡＰＩ

１３４　回復指示用ＡＰＩ

１４０　メモリ

２００　ディスク制御装置

２１０　チャネルポート

２２０　マイクロプロセッサ

２３０　バッファメモリ

３００　ディスク制御プログラム

３１０　書込み制御処理

３２０　書込みデータ処理

３３０　ディスク管理処理

３４０　データ回復制御処理

３５０　データ回復処理

３６０　データ同期処理

４００　大容量記憶装置

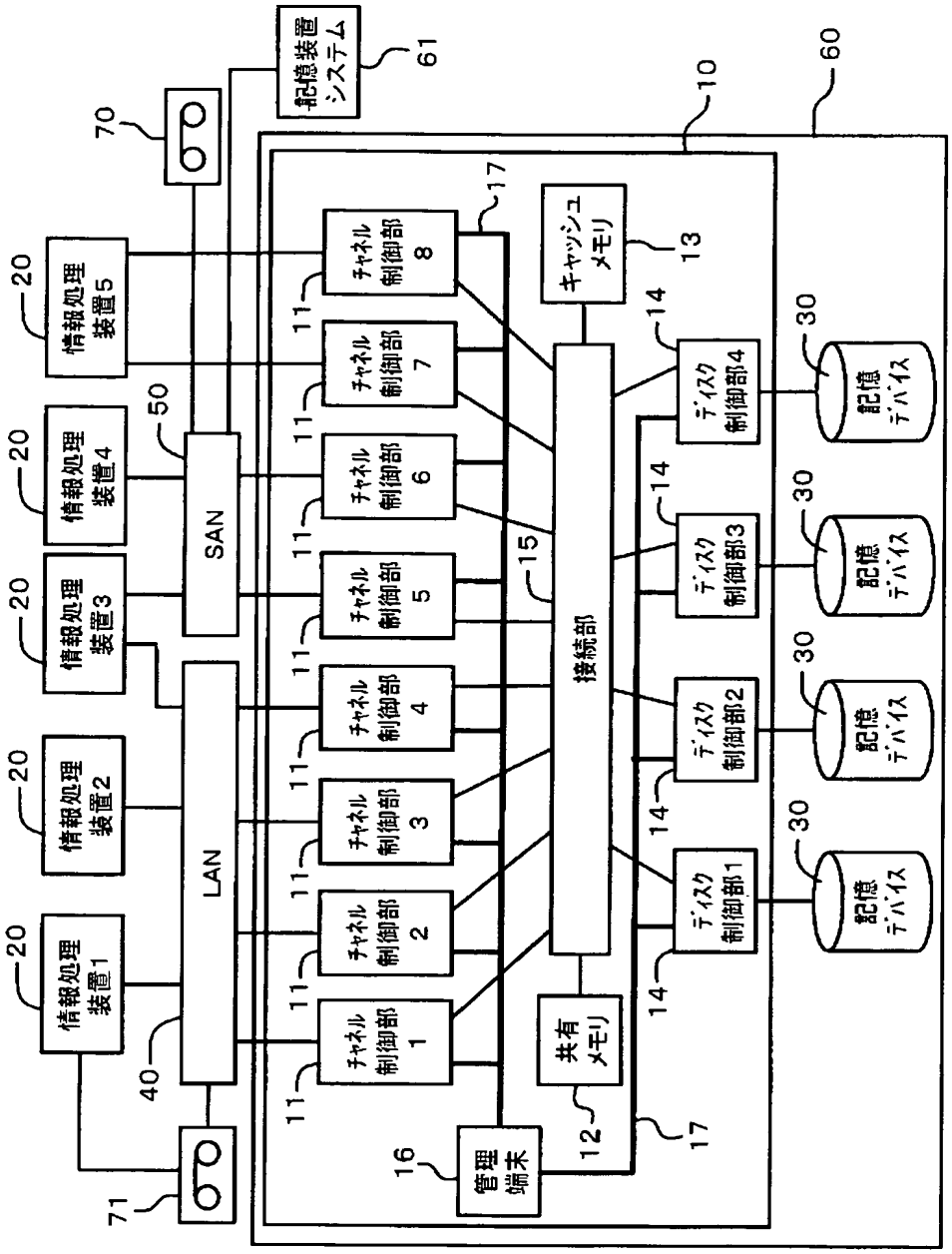
４１０　データ格納ディスク装置

４２０　バックアップデータ格納ディスク装置

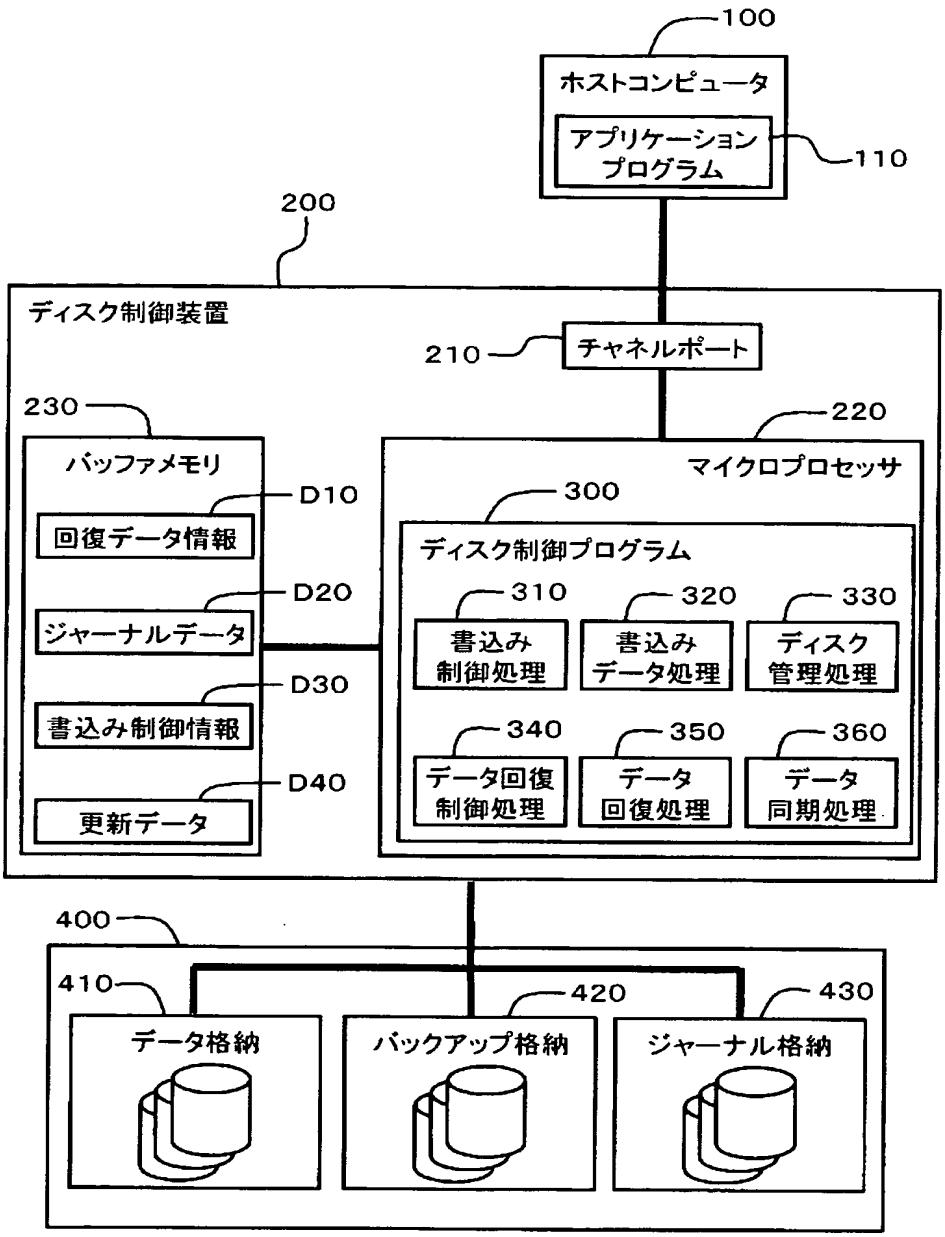
４３０　ジャーナルデータ格納ディスク装置

【書類名】　図面

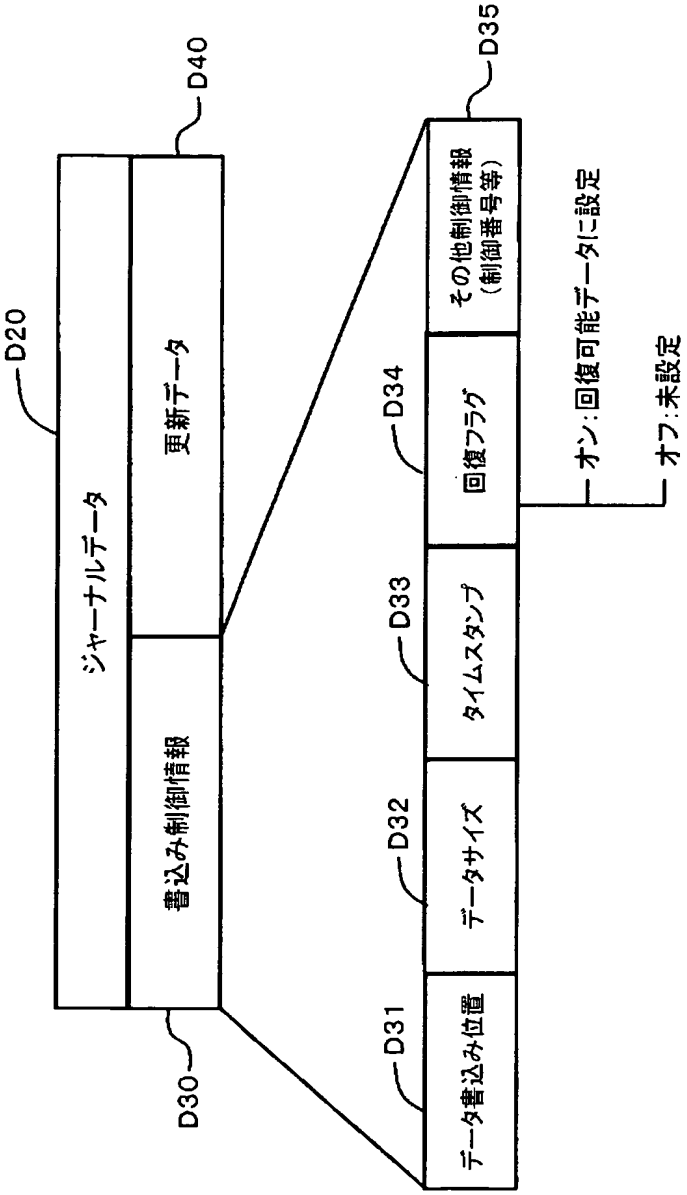
【図１】


記憶装置システム
61
60
10
70
17
チャネル制御部8
キャッシュメモリ
13
情報処理装置5
20
11
チャネル制御部7
14
ディスク制御部4
30
記憶デバイス
情報処理装置4
50
SAN
チャネル制御部6
15
接続部
情報処理装置3
チャネル制御部5
ディスク制御部3
情報処理装置2
LAN
チャネル制御部4
ディスク制御部2
チャネル制御部3
情報処理装置1
40
チャネル制御部2
ディスク制御部1
チャネル制御部1
共有メモリ
12
16
管理端末
71

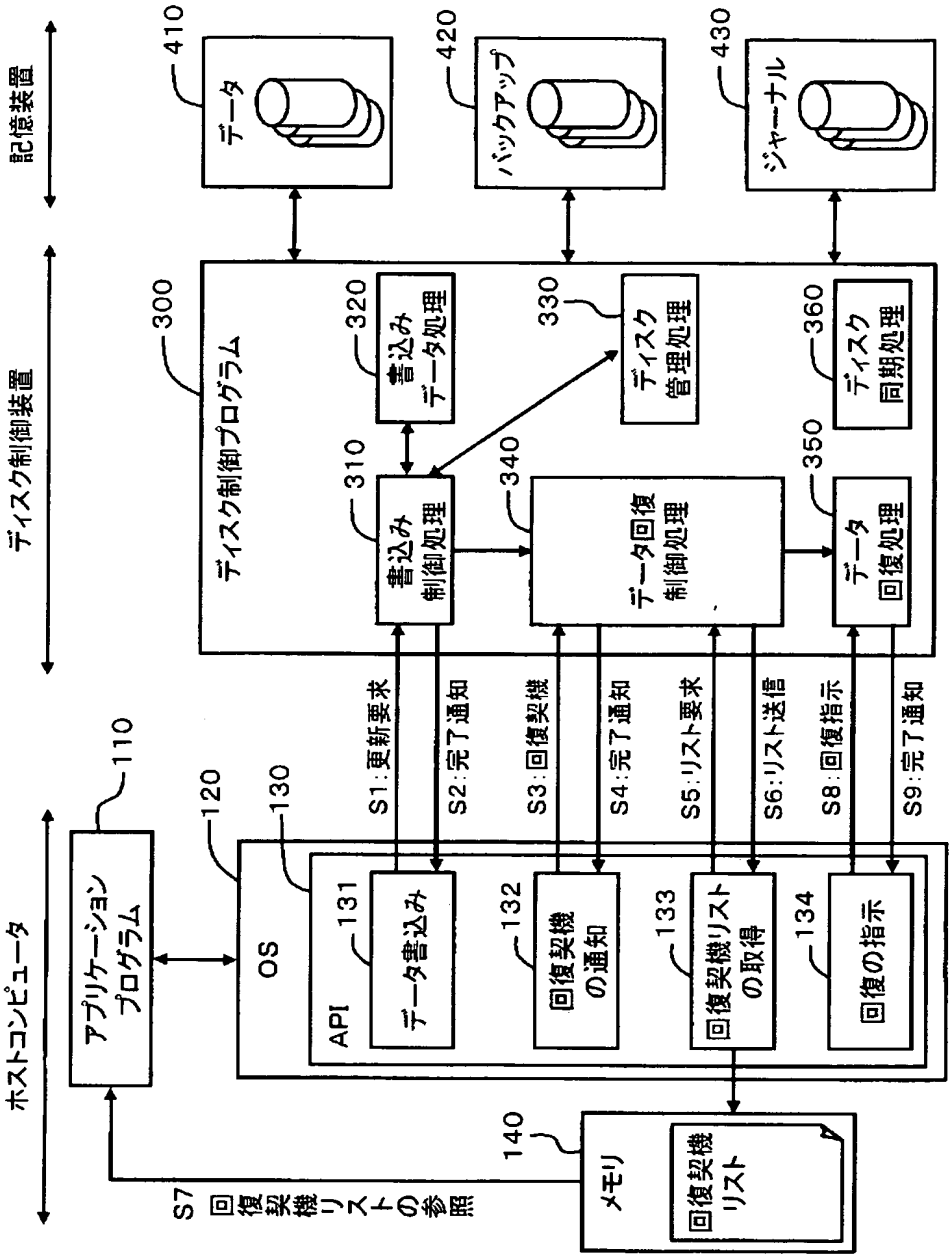
【図２】


100
ホストコンピュータ
アプリケーション
プログラム
110
200
ディスク制御装置
210
チャネルポート
230
バッファメモリ
D10
回復データ情報
D20
ジャーナルデータ
D30
書込み制御情報
D40
更新データ
220
マイクロプロセッサ
300
ディスク制御プログラム
310
書込み
制御処理
320
書込み
データ処理
330
ディスク
管理処理
340
データ回復
制御処理
350
データ
回復処理
360
データ
同期処理
400
410
データ格納
420
バックアップ格納
430
ジャーナル格納

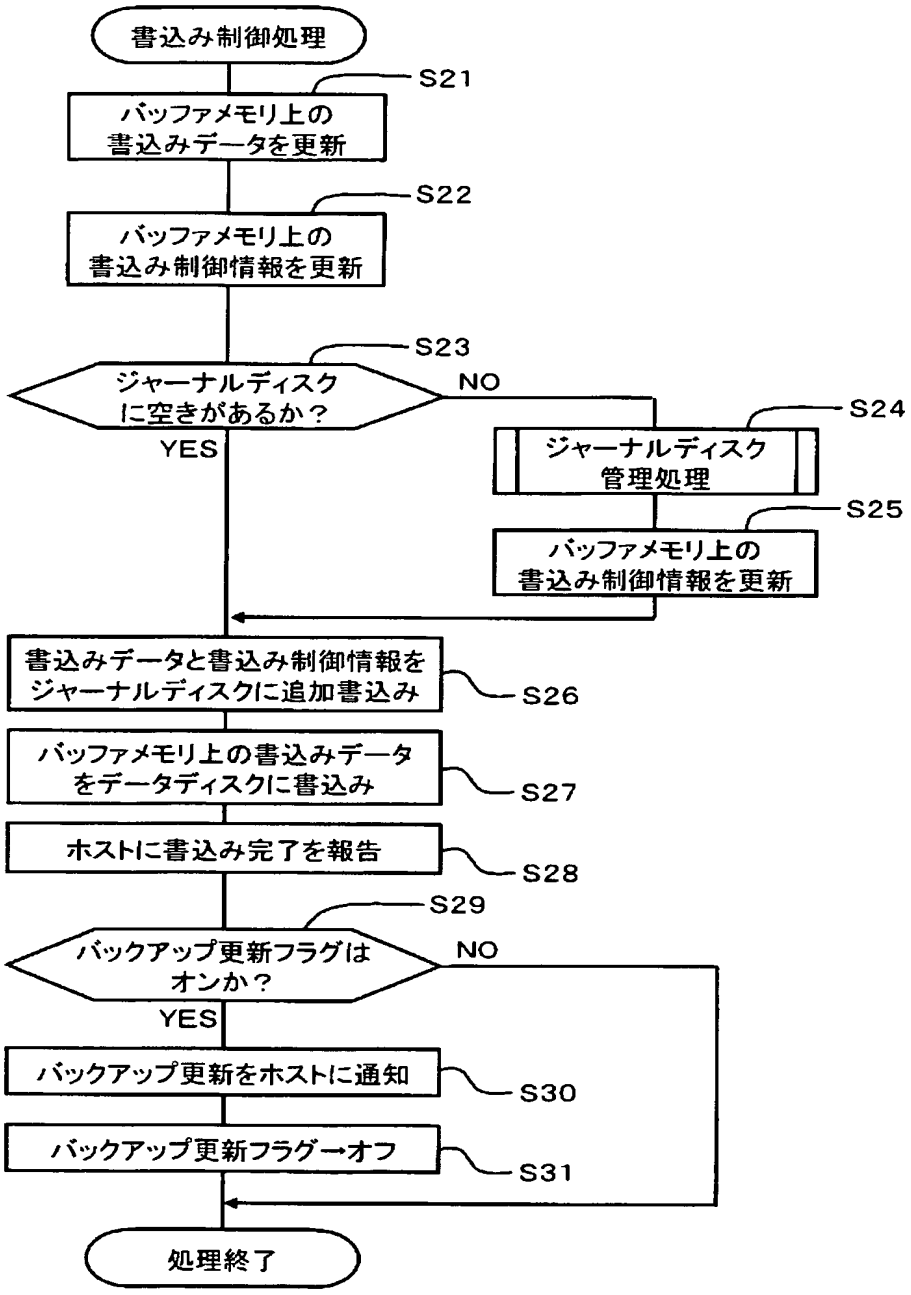
【図３】


D20
ジャーナルデータ
D40
更新データ
書込み制御情報
D30
D31
データ書込み位置
D32
データサイズ
D33
タイムスタンプ
D34
回復フラグ
オン：回復可能データに設定
オフ：未設定
D35
その他制御情報
（制御番号等）

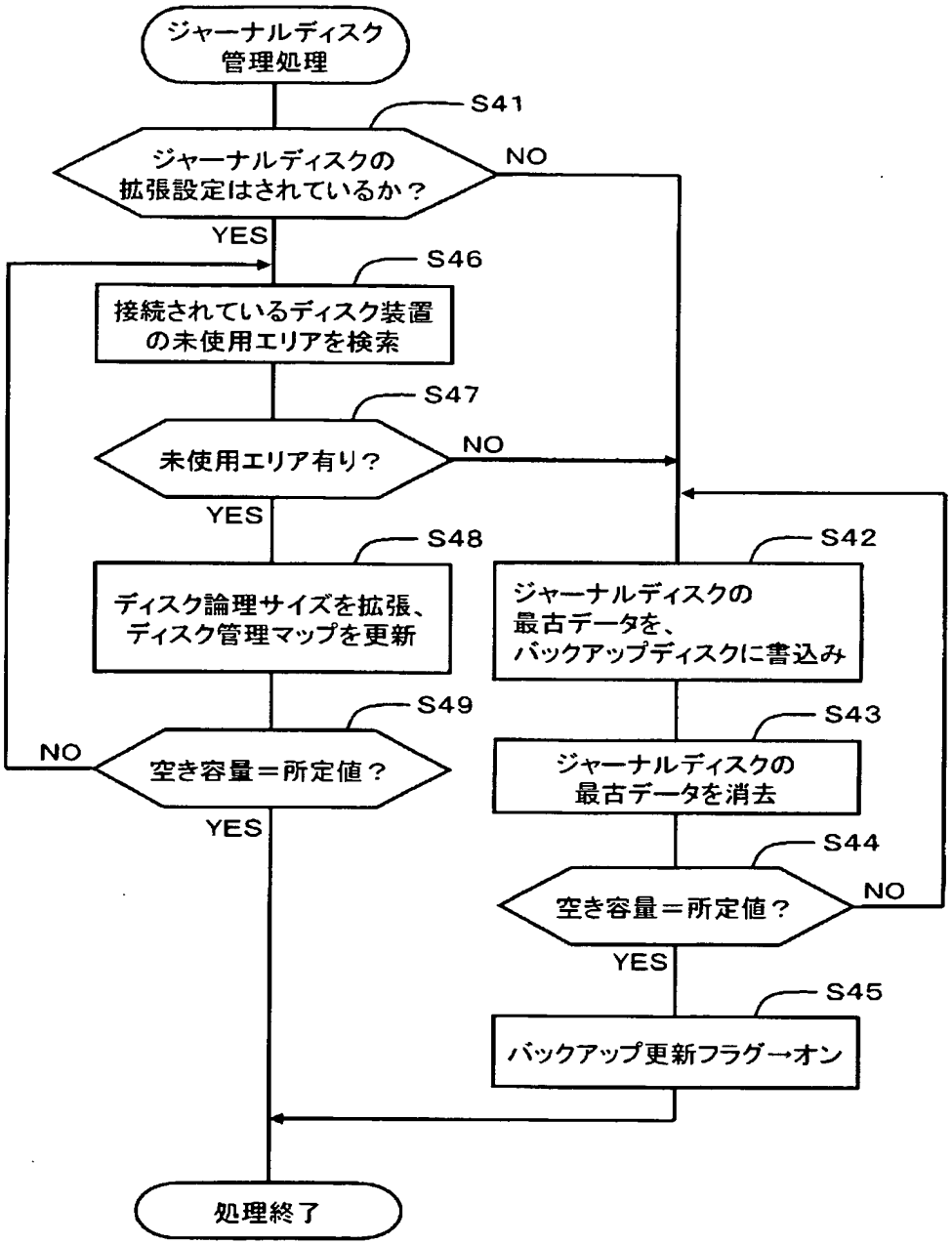
【図４】


記憶装置
ディスク制御装置
ホストコンピュータ
410
データ
420
バックアップ
430
ジャーナル
300
ディスク制御プログラム
320
書込み
データ処理
330
ディスク
管理処理
360
ディスク
同期処理
310
書込み
制御処理
340
データ回復
制御処理
350
データ
回復処理
110
アプリケーション
プログラム
120
130
OS
API
131
データ書込み
132
回復契機
の通知
133
回復契機リスト
の取得
134
回復の指示
140
メモリ
回復契機
リスト
S1:更新要求
S2:完了通知
S3:回復契機
S4:完了通知
S5:リスト要求
S6:リスト送信
S8:回復指示
S9:完了通知
S7
回復契機リストの参照

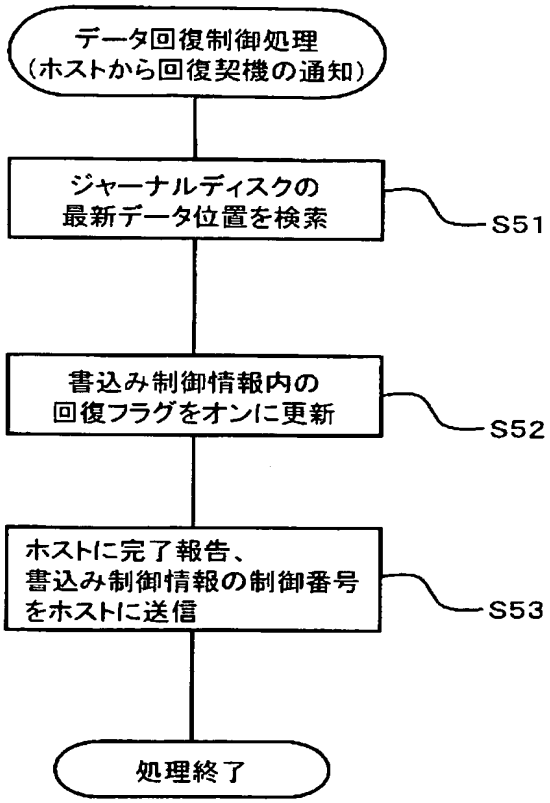
【図５】


書込み制御処理
S21
バッファメモリ上の
書込みデータを更新
S22
バッファメモリ上の
書込み制御情報を更新
S23
ジャーナルディスク
に空きがあるか？
NO
YES
S24
ジャーナルディスク
管理処理
S25
バッファメモリ上の
書込み制御情報を更新
書込みデータと書込み制御情報を
ジャーナルディスクに追加書込み
S26
バッファメモリ上の書込みデータ
をデータディスクに書込み
S27
ホストに書込み完了を報告
S28
S29
バックアップ更新フラグは
オンか？
NO
YES
バックアップ更新をホストに通知
S30
バックアップ更新フラグ→オフ
S31
処理終了

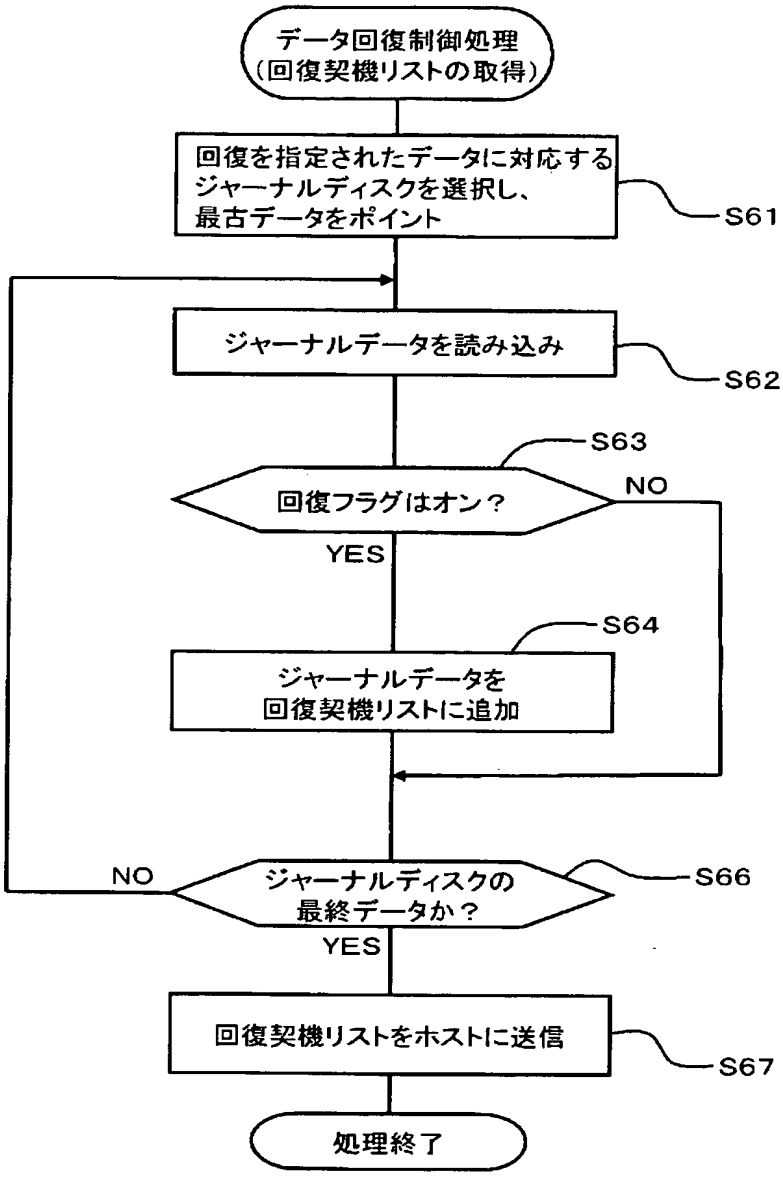
【図６】


ジャーナルディスク
管理処理
S41
ジャーナルディスクの
拡張設定はされているか？
NO
YES
S46
接続されているディスク装置
の未使用エリアを検索
S47
未使用エリア有り？
NO
YES
S48
ディスク論理サイズを拡張、
ディスク管理マップを更新
S49
NO
空き容量＝所定値？
YES
S42
ジャーナルディスクの
最古データを、
バックアップディスクに書込み
S43
ジャーナルディスクの
最古データを消去
S44
空き容量＝所定値？
NO
YES
S45
バックアップ更新フラグ→オン
処理終了

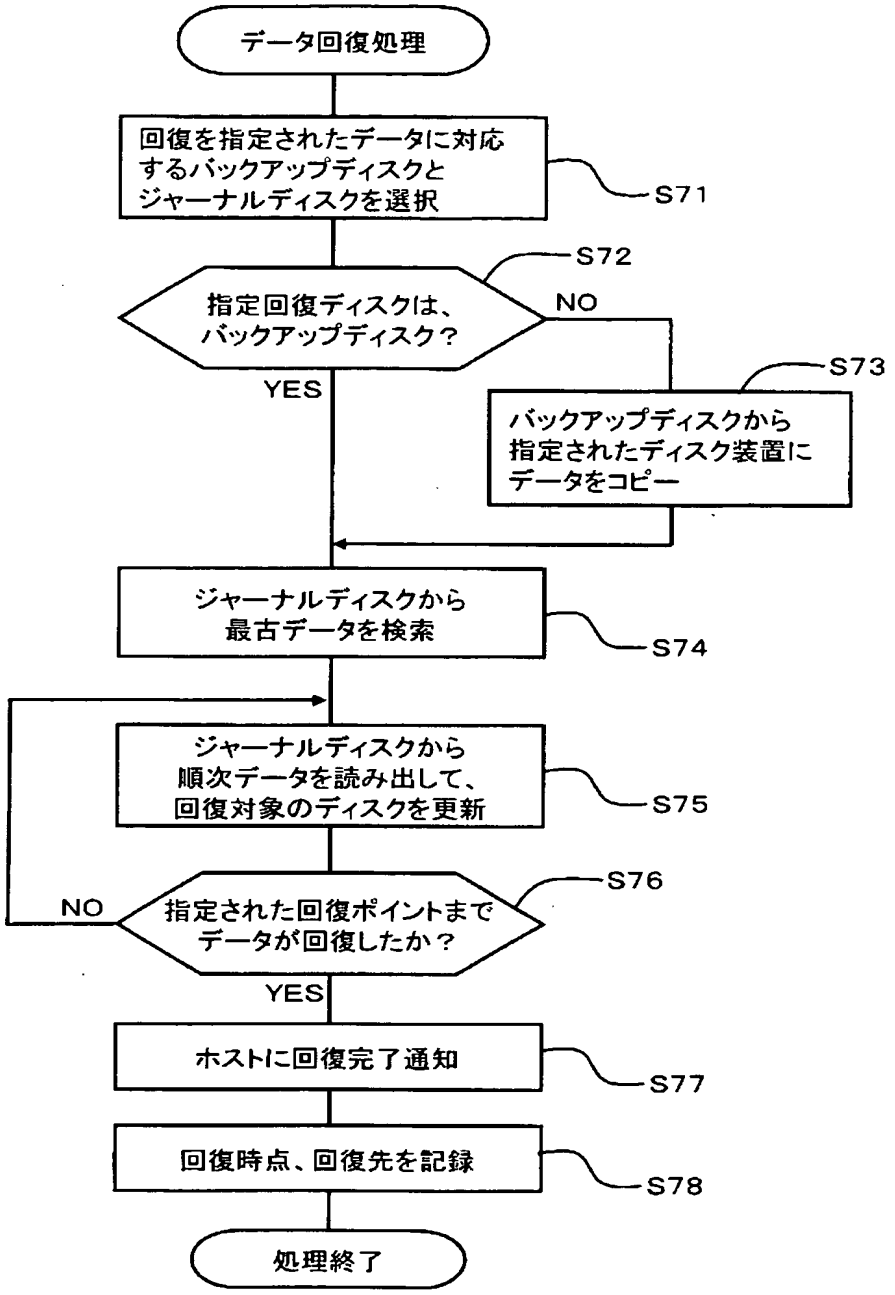
【図７】


データ回復制御処理
（ホストから回復契機の通知）
ジャーナルディスクの
最新データ位置を検索
S51
書込み制御情報内の
回復フラグをオンに更新
S52
ホストに完了報告、
書込み制御情報の制御番号
をホストに送信
S53
処理終了

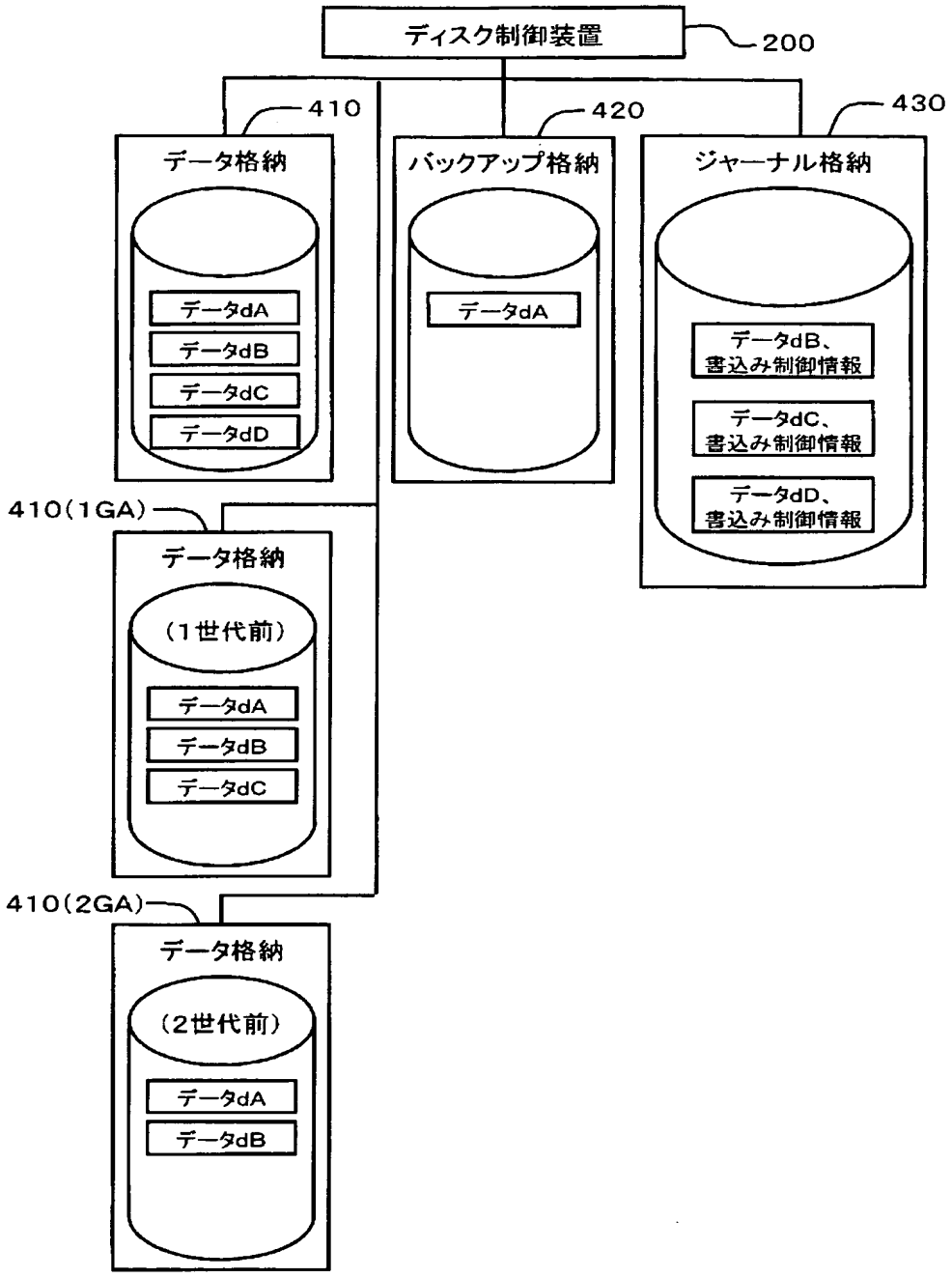
【図８】


データ回復制御処理
(回復契機リストの取得)
回復を指定されたデータに対応する
ジャーナルディスクを選択し、
最古データをポイント
S61
ジャーナルデータを読み込み
S62
S63
回復フラグはオン？
NO
YES
S64
ジャーナルデータを
回復契機リストに追加
NO
ジャーナルディスクの
最終データか？
S66
YES
回復契機リストをホストに送信
S67
処理終了

【図９】


データ回復処理
回復を指定されたデータに対応するバックアップディスクとジャーナルディスクを選択
S71
S72
指定回復ディスクは、バックアップディスク？
NO
YES
S73
バックアップディスクから指定されたディスク装置にデータをコピー
ジャーナルディスクから最古データを検索
S74
ジャーナルディスクから順次データを読み出して、回復対象のディスクを更新
S75
S76
NO
指定された回復ポイントまでデータが回復したか？
YES
ホストに回復完了通知
S77
回復時点、回復先を記録
S78
処理終了

【図１０】


ディスク制御装置
200
410
420
430
データ格納
バックアップ格納
ジャーナル格納
データdA
データdB
データdC
データdD
データdA
データdB、
書込み制御情報
データdC、
書込み制御情報
データdD、
書込み制御情報
410(1GA)
データ格納
（1世代前）
データdA
データdB
データdC
410(2GA)
データ格納
（2世代前）
データdA
データdB

【書類名】　　要約書

【要約】

【課題】　ホストコンピュータに負担をかけずに、外部記憶装置内で、所望の任意の時点までデータを自動的に回復させる。

【解決手段】　ホストコンピュータ上のアプリケーション１１０は、ディスク制御装置のデータ回復制御処理３４０に対し、回復契機の設定を指示する（Ｓ３）。ジャーナルデータ中に含ませた回復フラグをセットすることにより、任意の複数時点を回復可能な時点として登録させることが可能となっている。障害等が発生してデータを回復させる場合、アプリケーション１１０は、設定済の回復契機の一覧を示すリストを要求する（Ｓ５）。アプリケーション１１０は、回復契機リストに基づいて、データを回復させる時点を指定する（Ｓ８）。ディスク制御装置は、バックアップディスク４２０及びジャーナルディスク４３０に基づいて、指示された時点までデータを回復させる。

【選択図】　　図４

認定・付加情報

特許出願の番号	特願２００３－０７６８６５
受付番号	５０３００４５５３１７
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１５年　３月２４日

＜認定情報・付加情報＞

【提出日】　平成15年 3月20日

次頁無

特願２００３－０７６８６５

出　願　人　履　歴　情　報

識別番号　［０００００５１０８］

１．変更年月日　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所